CONNECTICUT
LAW REVIEW

VOLUME 52 | FEBRUARY 2021 | NUMBER 3

Article

Political Polarization and Moral Outrage on Social Media

JORDAN CARPENTER, WILLIAM BRADY, MOLLY CROCKETT, RENÉ WEBER & WALTER SINNOTT-ARMSTRONG

Many theorists claim that social media contribute to political polarization, but it is not clear how these effects occur. We propose and explain a theoretical model of this process that focuses on moral outrage. This combination of anger and disgust can emerge from a mismatch between evolved human nature and certain features of political discussions on the internet. We identify three specific types of socially negative behavior that moral outrage facilitates: aggression (behavior intended to harm others), sophistry (poor argumentation), and withdrawal (avoiding discussions of politics). We describe psychological mechanisms through which moral outrage can lead to these outcomes, specifically focusing on dehumanization and group antagonism. We discuss research justifying our proposed model and suggest new ways to empirically test its links. Our model should be useful for researchers exploring the question of when and how political discussions on social media go wrong as well as what to do about these problems.

ARTICLE CONTENTS

Political Polarization and Moral Outrage on Social Media

JORDAN CARPENTER, WILLIAM BRADY, MOLLY CROCKETT, RENÉ WEBER & WALTER SINNOTT-ARMSTRONG [*]

BACKGROUND

Decades ago, experts hailed the internet as a grand, new opportunity for political enlightenment. It was thought that the web would provide a convenient and widely available way to remove limitations imposed by geography and resources, expanding access to information, increasing understanding and empathy among people, and making the world better. Today, this optimistic view is tempered by fears that certain aspects of internet use—most notably social media—have the potential to exacerbate threats to democracy, including political polarization.[1]

Political polarization is sometimes understood merely as ideological distance between political parties or homogeneity within parties.[2] However, group coherence and disagreement by themselves are not the main problems here. The more threatening kind of polarization, which is often described as affective group polarization, involves intense, negative attitudes toward the political outgroup.[3] According to Pew Research

---

[*] William Brady is an NSF postdoctoral fellow in the psychology department at Yale University. Jordan Carpenter is a postdoctoral fellow in the Kenan Institute for Ethics at Duke University. Molly Crockett is an Assistant Professor of Psychology at Yale University and a Distinguished Research Fellow at the Oxford Centre for Neuroethics, University of Oxford. Walter Sinnott-Armstrong is the Chauncey Stillman Professor of Practical Ethics at Duke University in the Philosophy Department, the Kenan Institute for Ethics, the Psychology and Neuroscience Department, and the Law School. René Weber (M.D., University of Aachen, Germany; Ph.D. University of Technology Berlin, Germany) is a professor in the Department of Communication at the University of California in Santa Barbara, director of UCSB's Media Neuroscience Lab (https://medianeuroscience.org), and a Fellow of the International Communication Association.

[1] William J. Brady & M.J. Crockett, *How Effective Is Online Outrage?*, 23 TRENDS COGNITIVE SCI. 79, 79–80 (2019). *See* CASS R. SUNSTEIN, #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA 5–6 (2017) (advocating for "an architecture of serendipity" as the ultimate way to salvage democracy, as "[t]o the extent that social media allow us to create our very own feeds, and essentially live in them, they create serious problems. . . . Self-insulation and personalization . . . spread falsehoods, and promote polarization and fragmentation"); ANAMITRA DEB ET AL., IS SOCIAL MEDIA A THREAT TO DEMOCRACY? 3–4 (2017), https://www.omidyargroup.com/wp-content/uploads/2017/10/Social-Media-and-Democracy-October-5-2017.pdf (reporting the six key features of social media that challenge democratic principles).

[2] Christopher Hare & Keith T. Poole, *The Polarization of Contemporary American Politics*, 46 POLITY 411, 412 (2014).

[3] Shanto Iyengar & Sean J. Westwood, *Fear and Loathing Across Party Lines: New Evidence on*

surveys in 2014, deep antipathy toward one's political outgroup grew by 24% in the preceding decade among the American public, and nearly 32% of Americans saw the opposing party's policies as threats to the nation or its well-being.[4] These strong feelings have contributed to violent clashes, such as those between far-right political groups and liberals in New Orleans and Charlottesville.[5] More generally, increasing affective group polarization has led to a decline in the kind of civil discourse that many hold to be a cornerstone of democracy.[6]

While social media use is widely believed to contribute to growing polarization,[7] data directly addressing this claim are scarce and in part lead to controversial interpretations and conclusions. As a result, the processes through which social media might exacerbate polarization are not well understood. We need to figure out the processes behind polarization in order to figure out what to do about it. Solutions require understanding.

I. THESIS

We propose here that moral outrage is central to understanding how social media use is related to affective group polarization. Moral outrage is an intense negative emotion combining anger and disgust triggered by a perception that someone violated a moral norm.[8] Messages that describe or evoke moral outrage are increasingly prevalent in contemporary political contexts, especially those accusing political opponents of moral norm

---

Group Polarization, 59 AM. J. POL. SCI. 690, 690 (2015); Matt Motyl, *Liberals and Conservatives Are (Geographically) Dividing*, *in* SOCIAL PSYCHOLOGY OF POLITICAL POLARIZATION 7, 21 (Piercarlo Valdesolo & Jesse Graham eds., 2016).

[4] *Political Polarization in the American Public: Section 2: Growing Partisan Antipathy*, PEW RES. CTR. (June 12, 2014), https://www.people-press.org/2014/06/12/section-2-growing-partisan-antipathy/.

[5] *See* Nicholas Bogel-Burroughs, *What Is Antifa? Explaining the Movement to Confront the Far Right*, N.Y. TIMES (July 2, 2019), https://www.nytimes.com/2019/07/02/us/what-is-antifa.html (reporting on the movements of "antifa," a contraction of the word "anti-fascist," including protests in Charlottesville that turned violent); Alan Feuer & Jeremy W. Peters, *Fringe Groups Revel as Protests Turn Violent*, N.Y. TIMES (June 2, 2017), https://www.nytimes.com/2017/06/02/us/politics/white-nationalists-alt-knights-protests-colleges.html (describing various groups' attempts to mobilize, including that of the Proud Boys—a clan of conservative nationalists—in New Orleans over the removal of Confederate monuments).

[6] Jürgen Habermas, *Three Normative Models of Democracy*, 1 CONSTELLATIONS 1, 7 (1994); WALTER SINNOTT-ARMSTRONG, THINK AGAIN: HOW TO REASON AND ARGUE 2–4 (2018) [hereinafter, SINNOTT-ARMSTRONG, THINK AGAIN].

[7] Levi Boxell et al., *Greater Internet Use Is Not Associated with Faster Growth in Political Polarization Among US Demographic Groups*, 114 PNAS 10,612, 10,612–16 (2017); JOSHUA A. TUCKER ET AL., SOCIAL MEDIA, POLITICAL POLARIZATION, AND POLITICAL DISINFORMATION: A REVIEW OF THE SCIENTIFIC LITERATURE 3–5 (2018), https://hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf.

[8] Jessica M. Salerno & Liana C. Peter-Hagene, *The Interactive Effect of Anger and Disgust on Moral Outrage and Judgments*, 24 PSYCHOL. SCI. 2069, 2074 (2013).

violations.[9] The moral nature of such messages makes them more likely to capture audiences' attention[10] and intensifies receivers' emotional involvement.[11] The resulting moral outrage is associated with especially stubborn political views[12] and can even facilitate political violence.[13]

Recent theorizing suggests that the design of social media platforms amplifies moral outrage by lowering the social costs associated with outrage and increasing its personal benefits,[14] especially when moral content interacts with moral sensitivities to shape exposure to social media and subsequent behavior.[15] Thus, moral outrage sparked by messages on social media and the internet more broadly is likely a crucial factor in explaining recent alarming trends in societal discourse and their consequences for increasing polarization and the decay of democratic norms.

II. MODEL

To understand affective group polarization, we propose a model describing how a mismatch between our evolutionary past and current social media amplifies moral outrage in online contexts. This, among other factors, leads to affective group polarization, involving group antagonism and dehumanization, which subsequently motivates social behaviors that directly threaten democracy.

---

[9] Spassena P. Koleva et al., *Tracing the Threads: How Five Moral Concerns (Especially Purity) Help Explain Culture War Attitudes*, 46 J. RES. PERSONALITY 184, 191–93 (2012).

[10] William J. Brady et al., *Attentional Capture Helps Explain Why Moral and Emotional Content Go Viral*, J. EXPERIMENTAL PSYCHOL. GEN. 1, 4 (2019); Ana P. Gantman & Jay J. Van Bavel, *The Moral Pop-Out Effect: Enhanced Perceptual Awareness of Morally Relevant Stimuli*, 132 COGNITION 22, 28 (2014).

[11] William J. Brady et al., *Emotion Shapes the Diffusion of Moralized Content in Social Networks*, 114 PNAS 7313, 7316 (2017).

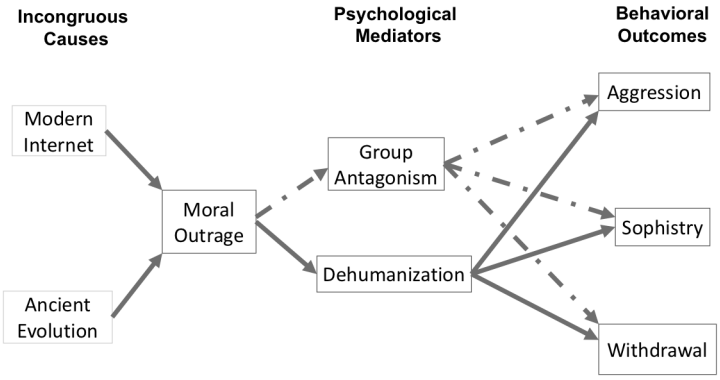
[12] Linda J. Skitka et al., *Moral Conviction: Another Contributor to Attitude Strength or Something More?*, 88 J. PERSONALITY & SOC. PSYCHOL. 895, 903 (2005) (testing "the degree that moral conviction was correlated" with political orientation).

[13] ALAN PAGE FISKE & TAGE SHAKTI RAI, VIRTUOUS VIOLENCE: HURTING AND KILLING TO CREATE, SUSTAIN, END, AND HONOR SOCIAL RELATIONSHIPS 1–2 (2015) (discussing virtuous violence theory); Marlon Mooijman et al., *Moralization in Social Networks and the Emergence of Violence During Protests*, 2 NATURE HUM. BEHAV. 389, 389 (2018).

[14] M. J. Crockett, *Moral Outrage in the Digital Age*, 1 NATURE HUM. BEHAV. 769, 769–71 (2017).

[15] Richard Huskey et al., *Things We Know About Media and Morality*, 2 NATURE HUM. BEHAV. 315, 315 (2018).

Figure 1: Proposed model of online moral outrage.



Our model begins with some fundamental sources of online moral outrage. Human psychology developed over evolutionary time in small communities, where observing egregious acts was a rare and noteworthy event. By contrast, the modern world, particularly with the development of social media, supplies a near-constant barrage of material that evokes moral outrage when political discussions occur. Other features of online contexts that exacerbate moral outrage include the psychological distance between conversation partners and the rarity of punitive consequences for bad behavior,[16] as well as the predominantly written nature of online communication, which can intensify the emotional impact of messages.[17] It was also much harder and more dangerous to leave one's small community in evolutionary times than it is to drop out of online exchanges. This mismatch between the circumstances in which our ancestors evolved and the online worlds that many of us inhabit today plays a large role in instigating the problem of moral outrage online.

In the next stage of our model, online moral outrage leads to two psychological states that characterize affective group polarization: *group antagonism* (antipathy toward groups of political opponents)[18] and *dehumanization* (failure to recognize others' human mental attributes).[19] These psychological states then lead to three distinct social behaviors: *aggression* (behavior intended to harm another individual),[20] *sophistry*

---

[16] Crockett, *supra* note 14, at 769–71.

[17] Huskey et al., *supra* note 15, at 315; Joseph B. Walther, *Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction*, 23 COMM. RES. 3, 3–5, 7–8 (1996).

[18] Shanto Iyengar & Sean J. Westwood, *Fear and Loathing Across Party Lines: New Evidence on Group Polarization*, 59 AM. J. POL. SCI. 690, 690, 704 (2015).

[19] Lasana T. Harris & Susan T. Fiske, *Dehumanizing the Lowest of the Low: Neuroimaging Responses to Extreme Out-Groups*, 17 PSYCHOL. SCI. 847, 847–48, 850 (2006); Nick Haslam, *Dehumanization: An Integrative Review*, 10 PERSONALITY & SOC. PSYCHOL. REV. 252, 252–53 (2006).

[20] ROBERT A. BARON & DEBORAH R. RICHARDSON, HUMAN AGGRESSION (PERSPECTIVES IN SOCIAL PSYCHOLOGY) 7 (2d ed. 1994).

(using empty, misleading, or irrelevant arguments),[21] and *withdrawal* (deliberately avoiding political participation, including voting, contribution, discussion, or even learning about political issues). These behaviors can threaten democracy by restricting communication, cooperation, civic participation, and the ability to react appropriately to political events.

III. LINKS

In order to test each link in this model, we need to pose a variety of research questions. We cannot answer any of these questions yet, but asking them will illuminate the central claims in our model and will show why we think our model is at least plausible.

A. *Does moral content increase moral outrage?*

Because moral outrage is triggered when a perceiver of a message believes an important moral norm has been violated, messages (e.g., tweets) without moral information are less likely to elicit moral outrage than messages that contain information about moral wrongdoing or moral conflict. In addition, the model of intuitive morality and exemplars (MIME)[22] has shown that effects of social media messages are intensified when their content addresses violations or upholdings of moral norms that the audience endorses and sees as important.[23] Furthermore, evidence from communication diffusion models repeatedly suggests that media effects are a function of both stimulus prevalence and stimulus density over a given time interval (e.g., the number of communicators or the number of message repetitions).[24] Hence, the high prevalence and density of moral information and moral conflict in social media could help to explain why social media trigger such strong emotions.

---

[21] *See* SINNOTT-ARMSTRONG, THINK AGAIN, *supra* note 6, at 183–84 (discussing the sophistical fallacy of misleading others by jumping topics and avoiding the question asked).

[22] Ron Tamborini, *Model of Intuitive Morality and Exemplars*, *in* MEDIA AND THE MORAL MIND 43, 43 (Ron Tamborini ed., 2015); Ron Tamborini & René Weber, *Advancing the Model of Intuitive Morality and Exemplars*, *in* THE HANDBOOK OF COMMUNICATION SCIENCE AND BIOLOGY 456, 456 (Kory Floyd & René Weber eds., 2020).

[23] Tamborini, *Model of Intuitive Morality and Exemplars*, *supra* note 22, at 50–51; *see also* Graham J. Haidt et al., *Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism*, *in* 47 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 55, 82, 83 (James Olson ed., 2013) (discussing a study that showed individuals "were more likely . . . to favor those who personified virtues related to" ideals that are stereotypically valued by the side of the political spectrum with which the individual identified).

[24] *See* Ronald E. Rice, *Intermediality and the Diffusion of Innovations*, 43 HUM. COMM. RES. 531, 531 (2017) (discussing the communication diffusion perspective of innovation).

B. *Do social media amplify moral outrage?*

Moral outrage has been singled out as especially likely to occur in online political discourse.[25] Although moral outrage can also occur in face-to-face interactions, several factors exacerbate its effects online. Offline, people rarely encounter egregious moral violations, but social media and other technologies allow people to become aware of others' worst behaviors much more easily. People are highly motivated to express outrage about immoral actions, which makes such information especially likely to go viral.[26] In addition, expressing outrage is easy online, because the target of the outrage need not be present, the potential for retaliation is minimal, and distant targets inspire less empathic concern.[27] It is also easy to express outrage and leave the website without waiting for any response and, hence, without knowledge of how harmful one's expression might have been.

Data from previous studies using experience sampling[28] suggest that people experience more intense outrage in response to immoral events that they encounter online compared to events that they encounter in person or via traditional media (e.g., TV, radio, newspaper).[29] Spending time on social media would therefore seem to increase the likelihood of experiencing strong moral outrage.

These effects might be moderated by other factors, including age and political ideology. Age is of particular interest in light of recent evidence that older Americans have shown the greatest increases in group antagonism in recent years, despite using the internet and social media the least;[30] however, other evidence suggests that older adults who do use social media are the most polarized.[31] Certain demographic groups, such as older people, may be more vulnerable to the exacerbation of online moral outrage than others, which may help to explain demographic asymmetries in polarization. Just as not everyone exposed to a virus will fall ill, not everyone exposed to partisan content online will be influenced in the same

---

[25] *See* Crockett, *supra* note 14, at 769 (explaining how the internet facilitates the spread of moral outrage).

[26] Brady et al., *supra* note 11, at 7316.

[27] Crockett, *supra* note 14, at 770.

[28] MIHALY CSIKSZENTMIHALYI, FLOW AND THE FOUNDATIONS OF POSITIVE PSYCHOLOGY 21 (2014) ("The Experience Sampling Method (ESM) is a research procedure for studying what people do, feel, and think during their daily lives."). For examples of studies using experience sampling, see Crockett, *supra* note 14, at 770 and Wilhelm Hofmann et al., *Morality in Everyday Life*, 345 SCIENCE 1340, 1340–41 (2014).

[29] Crockett, *supra* note 14, at 770.

[30] Boxell et al., *supra* note 7, at 10,612.

[31] *Id.*; *see also National Politics on Twitter: Small Share of U.S. Adults Produce Majority of Tweets*, PEW RES. CTR. (Oct. 23, 2019), https://www.people-press.org/2019/10/23/national-politics-on-twitter-small-share-of-u-s-adults-produce-majority-of-tweets/ (concluding that older Americans are tweeting the most about national politics).

way. Because older adults show changes in brain systems related to processing social feedback,[32] older adults could be differentially susceptible to online amplification of moral outrage.

The specific mechanisms by which internet usage combines with evolved characteristics can be further specified. For instance, it is well documented in the literature that repeated exposure to media, including social media, influences emotions and behaviors by altering the *salience* of moral and political content.[33] In a 2010 study by Leidner, Castano, Zaiser, and Giner-Sorolla, content that emphasized in-group glorification reduced the demands for justice when a violent perpetrator was an in-group member.[34] Moreover, this effect was mediated by moral disengagement (de-emphasizing suffering by victims' families and dehumanizing victims), which in turn is linked to violence and terrorism.[35]

The MIME mentioned above suggests that, over time, exposure to a consistent communication diet emphasizing the superiority of one moral intuition over another will either increase the salience of the emphasized intuitions or maintain their salience in the face of opposing influences. According to the MIME, polarization is expected in relatively closed systems, where outside influence is limited or blocked (such as in fundamentalist religious or political groups); whereas self-regulation is more likely in relatively open systems where external factors exert opposing forces (as in social media networks with fast and inexpensive information). The MIME holds that more isolated communicative networks with insulation from value-inconsistent messages should foster polarized values within such groups, intensify responses to moral conflicts between groups, and reduce openness to divergent views. Several studies have found these predicted effects in media content produced for and consumed by sub-groups that differ by age, political interest and orientation, moral intuition salience, culture, location, and dosage of exposure.[36]

---

[32] *See* Lars Bäckman et al., *The Correlative Triad Among Aging, Dopamine, and Cognition: Current Status and Future Prospects*, 30 NEUROSCIENCE & BIOBEHAVIORAL REV. 791, 797 (2006); Jean-Claude Dreher et al., *Age-Related Changes in Midbrain Dopaminergic Regulation of the Human Reward System*, 105 PNAS 15,106, 15,109 (2008); Ben Eppinger et al., *Reduced Striatal Responses to Reward Prediction Errors in Older Compared with Younger Adults*, 33 J. NEUROSCIENCE 9905, 9908 (2013); Shu-Chen Li et al., *Dopaminergic Modulation of Cognition Across the Life Span*, 34 NEUROSCIENCE & BIOBEHAVIORAL REV. 625, 628 (2010).

[33] William J. Brady, Killian McLoughlin & Molly J. Crockett, *Theory-Driven Measurement of Emotion (Expressions) in Social Media Text*, *in* THE ATLAS OF LANGUAGE ANALYSIS IN PSYCHOLOGY (Morteza Dehghani & Ryan Boyd eds.) (forthcoming) (manuscript at 17).

[34] Bernhard Leidner et al., *Ingroup Glorification, Moral Disengagement, and Justice in the Context of Collective Violence*, 36 PERSONALITY & SOC. PSYCHOL. BULL. 1115, 1116 (2010).

[35] *E.g.*, Alfred L. McAlister et al., *Mechanisms of Moral Disengagement in Support of Military Force: The Impact of September 11*, 25 J. SOC. & CLINICAL PSYCHOL. 141, 162–63 (2006).

[36] For an overview, see Tamborini & Weber, *supra* note 22, at 457–58.

C. *Does digital outrage increase group antagonism?*

Antagonism is not mere partisan disagreement. It involves hatred of political opponents in contrast to civil, substantive disputes about values and policies.[37] Our model concerns antagonism rather than civil disagreement.

Antagonism is related to moral outrage in that they both involve intensely negative emotions. However, whereas moral outrage is usually a response to an *individual* person or behavior, political antagonism is often directed against groups. Such political group antagonism is characterized by feelings of hostility towards the other political party and by beliefs that the other party is dangerous or evil. What began as a negative feeling towards an individual person or act grows into antagonism to their entire group.

D. *Does digital outrage lead to dehumanization?*

Dehumanization is a process of denying a person abilities and tendencies that are typical of human mental life.[38] It is distinct from antagonism in that it is possible to hate someone without dehumanizing them and vice versa.[39] However, we hypothesize that antagonism and dehumanization can feed one another and co-occur in the context of contentious political discourse online.

Dehumanization takes two distinct forms: a target can be denied *agency* (the ability to make reasonable decisions) or *feeling* (the ability to suffer).[40] People see the other side as "less than human"[41] either in their ability to reason or in their ability to feel pain. Both kinds of dehumanization can be a consequence of moral outrage, largely because of its emotional element of disgust, which is associated with

---

[37] Shanto Iyengar et al., *Affect, Not Ideology: A Social Identity Perspective on Polarization*, 76 PUB. OPINION Q. 405, 405, 408, 421 (2012).

[38] Nick Haslam, *Dehumanization: An Integrative Review*, 10 PERSONALITY & SOC. PSYCHOL. REV. 252, 252, 254 (2006); *see also* Lasana T. Harris & Susan T. Fiske, *Dehumanizing the Lowest of the Low: Neuroimaging Responses to Extreme Out-Groups*, 17 PSYCHOL. SCI. 847, 847 (2006) ("[W]e present new social neuroscience data indicating that extreme forms of prejudice may deny their targets even full humanity.").

[39] Tage S. Rai et al., *Dehumanization Increases Instrumental Violence, but Not Moral Violence*, 114 PNAS 8511, 8514–15 (2017).

[40] *See* Mengyao Li, Bernhard Leidner & Emanuele Castano, *Toward a Comprehensive Taxonomy of Dehumanization: Integrating Two Senses of Humanness, Mind Perception Theory, and Stereotype Content Model*, 21 TPM 285, 287 (2014) (defining "agency" as "the capacity for planning and acting" and defining "experience" as "the capacity for desires and feelings"). For further discussion of experience and agency, see Heather M. Gray et al., *Dimensions of Mind Perception*, 315 SCIENCE 619, 619 (2007).

[41] Madeleine Dalsklev & Jonas Rønningsdalen Kunst, *The Effect of Disgust-Eliciting Media Portrayals on Outgroup Dehumanization and Support of Deportation in a Norwegian Sample*, 47 INT'L J. INTERCULTURAL REL. 28, 29 (2015).

dehumanization.[42] Studies have also found that communication by text as opposed to voice leads to greater dehumanization,[43] so the fact that most online communication takes the form of writing might increase its contribution to dehumanization.

On the other hand, recent evidence suggests that people do not dehumanize the victims of violence when those victims are perceived as immoral, which is likely to be the case in the context of political conflict.[44] Perceiving someone as immoral in fact usually requires perceiving them as having nefarious or malicious intentions, which are human mental states. Merely humanizing opponents by ascribing some mental states to them is then not enough to forestall antagonism and aggression towards them.[45] Beliefs that they have bad intentions can instead make their suffering seem less aversive and increase antagonism and aggression towards them.[46] In this way, inaccurate perceptions of others' mental states can sometimes be just as pernicious as dehumanization.

The sources of group antagonism and dehumanization need to be determined in order to design remedies. Many proposed interventions on affective group polarization (such as those designed to increase empathy for political opponents) are predicated on the assumption that affective polarization leads people to spontaneously generate limited, simplistic theories about their opponents' motivations or emotions. These interventions are unlikely to succeed if their assumptions are inaccurate.[47]

E. *How can we test these links?*

To verify or falsify these assumptions, we need to measure relationships among outrage, group antagonism, and dehumanization among social media users. This task can now be approached with tools that have become available only recently, such as natural language processing[48]

---

[42] *Id.* at 29, 37–38; Katrina M. Fincher & Philip E. Tetlock, *Perceptual Dehumanization of Faces Is Activated by Norm Violations and Facilitates Norm Enforcement*, 145 J. EXPERIMENTAL PSYCHOL. 131, 132 (2016); Harris & Fiske, *supra* note 38, at 852.

[43] Juliana Schroeder et al., *The Humanizing Voice: Speech Reveals, and Text Conceals, a More Thoughtful Mind in the Midst of Disagreement*, 28 PSYCHOL. SCI. 1745, 1746, 1760 (2017).

[44] Rai et al., *supra* note 39, at 8513–14.

[45] *Id.* at 8512.

[46] *Id.* at 8511–12.

[47] *See* Scott Barry Kaufman, *Can Empathic Concern Actually Increase Political Polarization?*, SCI. AM. (Nov. 6, 2019), https://blogs.scientificamerican.com/beautiful-minds/can-empathic-concern-actually-increase-political-polarization/ (discussing how biases are likely to increase hostility toward the "outgroup").

[48] Frederic R. Hopp et al., *The Extended Moral Foundations Dictionary (eMFD): Development and Applications of a Crowd-Sourced Approach to Extracting Moral Intuitions from Text*, BEHAV. RES. METHODS 2 (2020), https://doi.org/10.3758/s13428-020-01433-0; CHRISTOPHER D. MANNING & HINRICH SCHÜTZE, FOUNDATIONS OF STATISTICAL NATURAL LANGUAGE PROCESSING 4 (1999); Eyal Sagi & Morteza Dehghani, *Measuring Moral Rhetoric in Text*, 32 SOC. SCI. COMPUTER REV. 132, 142

and supervised learning classification.[49] We predict that these tools can be used to uncover a positive relationship between expressions of moral outrage online and language that expresses antagonism towards groups and that dehumanizes opponents, such as by referring to them as animals.[50] We also predict that social reinforcement of expressions of moral outrage (in the form of "likes" and "retweets") will increase subsequent use of antagonistic and dehumanizing language in online discourse so that participants who receive the greatest amount of positive social feedback when they express moral outrage in their social media posts will show the highest levels of antagonism and dehumanization. Evidence for these predictions would support the corresponding links in our model between online moral outrage and the two psychological mediators: group antagonism and dehumanization.

F. *Does online moral outrage lead to aggression?*

Our model's next set of research questions asks whether moral outrage, through the mediators of both group antagonism and dehumanization, will lead to certain behaviors. Our model focuses on three actions: aggression, sophistry, and withdrawal.

To understand online *aggression*, recall that moral outrage begins as a negative emotional reaction to a single individual's act,[51] whereas antagonism is directed towards a group.[52] The transition from moral outrage to antagonism thus involves the spreading of negative feeling from one person to their entire group.

Anger at outgroups is associated with prejudice[53] and has been shown to be related specifically to disliking political outgroups more and tolerating them less.[54] Therefore, higher levels of moral outrage tend to lead to higher levels of prejudice and intolerance towards groups

---

(2013); René Weber et al., *Extracting Latent Moral Information from Text Narratives: Relevance, Challenges, and Solutions*, 12 COMM. METHODS & MEASURES 119, 124, 137 (2018).

[49] Brady, McLoughlin & Crockett, *supra* note 33 (manuscript at 6–9).

[50] *See, e.g.*, Florian Arendt & Narin Karadas, *Content Analysis of Mediated Associations: An Automated Text-Analytic Approach*, 11 COMM. METHODS & MEASURES 105, 112 (2017) (analyzing the use of animal-related terms over a four-month period to demonstrate the dehumanization of Muslims in German news coverage of Islam).

[51] *See* Salerno & Peter-Hagene, *supra* note 8, at 2069 (closely linking moral outrage with anger).

[52] *See* Giulia Evolvi, *#Islamexit: Inter-Group Antagonism on Twitter*, 22 INFO. COMM. & SOC'Y 386, 397 (2019) (studying group antagonism through anti-Muslim tweets during the Brexit debate in the United Kingdom).

[53] Nilanjana Dasgupta et al., *Fanning the Flames of Prejudice: The Influence of Specific Incidental Emotions on Implicit Prejudice*, 9 EMOTION 585, 589 (2009).

[54] Linda J. Skitka et al., *Political Tolerance and Coming to Psychological Closure Following the September 11, 2001, Terrorist Attacks: An Integrative Approach*, 30 PERSONALITY & SOC. PSYCHOL. BULL. 743, 754 (2004).

associated with the particular source of the moral outrage and then to representatives of those groups.

As a result, it seems likely that increased group antagonism will make people more willing to act aggressively to individual members of opposing groups based on group membership. Similarly, previous research has found that dehumanization is strongly associated with aggression, such that when a person perceives opponents as lacking feeling, they become more willing to inflict harm against opponents through bullying or harassment.[55] Therefore, dehumanization, particularly a lack of concern for the feelings of the target, would also seem to lead to aggression, at least in some cases.

G. *Does online moral outrage increase sophistry?*

Moral outrage also seems to lead people to engage in *sophistry*, or bad arguments, partly because one component of outrage is anger, which impairs judgment and decision making.[56] Ideally, the purpose of presenting arguments is to increase understanding of opposing points of view (including why others hold those positions) as well as to influence beliefs and attitudes on both sides of a controversy.[57] However, for many people talking about politics in social media, the focus is instead on competition and provocation (beating opponents by embarrassing, exhausting, or bewildering them) or theater (appearing more intelligent to observers who are allies or potential allies).[58] Even when people intend to create good arguments against opposing positions (or for their own), they often miss their targets because of a simplistic understanding of their opponents. This tendency seems to be exacerbated by higher levels of antagonism and dehumanization, which leads people to adopt a competitive or theatrical mindset during political discussions online, resulting in sophistry.

H. *Does online moral outrage motivate withdrawal?*

For the same reasons that moral outrage is galvanizing for some people, it leads others to *withdraw* from politics.[59] Intense animus can be overwhelming and unpleasant and will motivate many people to withdraw

---

[55] Albert Bandura et al., *Disinhibition of Aggression Through Diffusion of Responsibility and Dehumanization of Victims*, 9 J. RES. PERSONALITY 253, 266 (1975); Brock Bastian et al., *The Roles of Dehumanization and Moral Outrage in Retributive Justice*, 8 PLOS ONE 1, 9 (2013).

[56] Jennifer S. Lerner & Larissa Z. Tiedens, *Portrait of the Angry Decision Maker: How Appraisal Tendencies Shape Anger's Influence on Cognition*, 19 J. BEHAV. DECISION MAKING 115, 132 (2006).

[57] SINNOTT-ARMSTRONG, THINK AGAIN, *supra* note 6, at 56.

[58] *See, e.g.*, Ashley A. Anderson & Heidi E. Huntington, *Social Media, Science, and Attack Discourse: How Twitter Discussions of Climate Change Use Sarcasm and Incivility*, 39 SCI. COMM. 598, 600 (2017) (analyzing the use of sarcastic or uncivilized rhetoric in online discourse surrounding climate change).

[59] Elizabeth A. Bennett et al., *Disavowing Politics: Civic Engagement in an Era of Political Skepticism*, 119 AM. J. SOC. 518, 518–19 (2013).

in order to avoid or reduce associated negative emotions.[60] It is not clear what leads one person to become aggressive and another person to withdraw, but various individual differences might moderate these effects, such as the degree to which people experience moral outrage as unpleasant. In any case, antagonism and dehumanization by politically active people seem to lead some people into cynicism or apathy about politics. Just as many people effortfully avoid feeling sympathy for widespread suffering out of a desire to avoid emotional exhaustion,[61] so many people are motivated to avoid engaging with politics in order to keep from experiencing the hostility that characterizes contemporary partisan politics. They see politics as unpleasant, difficult, and exhausting; they foresee few compensating benefits for engaging in political activity, especially because of the sophistry and vicious attacks that characterize so much of political discourse online. For such reasons, both antagonism and dehumanization seem to lead many people to withdraw from politics—and understandably so.

IV. IMPACT

Many claims in our model remain speculations in need of further empirical support, but it could prove important and useful. If even approximately correct, our proposed model and its further specifications could illuminate the sources of many unpleasant psychological states and politically harmful behaviors on social media and elsewhere. It could help us understand an important social problem by providing a greater sense of the emotional and cognitive factors that lead people to behave badly when engaging in politics online. Because we need to understand a problem before we can solve it, our model could also potentially guide interventions that reduce political polarization and ensuing social problems.

All of this remains to be seen, because our model so far is only that: a hypothesized model—an educated guess. We would never claim to have established it as accurate. Much more research needs to be done to test it. All we can claim for now is that we find it plausible, promising, and potentially useful. We hope that others do, too.

---

[60] *Id.*

[61] C. Daryl Cameron & B. Keith Payne, *Escaping Affect: How Motivated Emotion Regulation Creates Insensitivity to Mass Suffering*, 100 J. PERSONALITY & SOC. PSYCHOL. 1, 2–3 (2011).